SYNOPSIS

ANALYSIS OF TAMIL STUTTERED SPEECH USING COMPUTATIONAL INTELLIGENCE

By

Ms. MANJUTHA M

16PHCSP004

Supervisor

Dr. P. Subashini

Professor

Department of Computer Science

Submitted to

School of Physical Science and Computational Science

Avinashilingam Institute for Home Science and Higher Education for Women

Coimbatore – 6410 043

AUGUST 2022

ANALYSIS OF TAMIL STUTTERED SPEECH USING COMPUTATIONAL INTELLIGENCE

1. INTRODUCTION

Speech disfluency is possibly a disabling situation with widespread and lifelong implications that may change individual behaviour, cognition, and social and emotional well-being. One speech-impaired disorder is stuttering, characterized by disruptions in the flow of speech or disfluencies. Stuttering people have some significant symptoms like repetitions of syllables, words, interjections, broken words, mispronounced words or syllables, vague phrases and certain prolonged sounds, which may severely affect the social development of stutter. People who stutter may be persistent, which is very challenging to treat at later ages.

In 2014, *The Stuttering Foundation* and Frank Wolf reported that worldwide more than 70 million people stutter, in which one in every 100 population has speech dysfluencies [15]. In 2010, "*Daily News and Analysis*" conveyed that India has 11 to 12 million people affected by stuttering, and 11 lakhs of people were adults [14]. *Census of India 2011* identifies that 21 lakhs have speech dysfluencies [4]. *The Indian Stammering Association (TISA)* supports 1.25 crore Indians who stammer, as inferred by "The Better India" in 2015 [13]. In 2017, *the National Medical Journal of India* identified that Tamil Nadu state holds many speech impairments people [3]. Statistical reports and literature reveal that many people were affected by stuttering. Hence, it is necessary to enhance the stuttered speech analysis, especially for the Tamil Language.

In the proposed research work, an active part of the speech signal is identified through Particle Swarm Optimization based Voice Activity Detection (PSO-VAD). Vocal tone and formant analysis are proposed to assess the speech fluency of Normal Articulate Speech (NAS), Moderate Stutter Speech (MSS) and Severe Stutter Speech (SSS). The effectiveness of speech disfluencies is computed by introducing regularity of speech energy and voice, which is further optimized with computational intelligence algorithms such as Particle Swarm Optimization (PSO), Synergistic Fibroblast Optimization (SFO) and Artificial Bee Colony (ABC). Hence the proposed research efficiently analyses the Tamil stuttered speech, which supports speech pathologists to diagnose the stuttered speech disorder more accurately.

2. DEFINITION OF THE PROBLEM

This research work proposes an efficient stuttered speech analysis that enhances speech disorders detection and diagnosis performance by introducing computational intelligence algorithms, especially for Tamil stuttered speech.

3. OBJECTIVES OF THE RESEARCH WORK

- To capture the characteristic features such as stop gaps, syllable repetition, vowel prolongation, Noisy Breathing, blocks, and interjection of stuttered speech.
- To analyse the Stutter Speech Severity Rate and Accuracy of Data Acquisition of the fluent and dysfluent speech signal.
- To detect and compare the utterance and severity Rate of Normal speech, Moderate and Severe Stutter speech.
- To detect and classify fluent and dysfluent speech into voiced, unvoiced, and silent speech signals.
- To detect and distinguish the vocal tone frequency variations and Formant frequency fluctuation among Normal Articulate Speech (NAS), Moderate Stutter Speech (MSS) and Severe Stutter Speech (SSS).
- To distinguish the speech regularity of fluent and dysfluent speech through energy, voice, or fundamental frequency.

4. METHODOLOGY

The process of proposed research methodology and the contribution of the research is shown in the following figure.

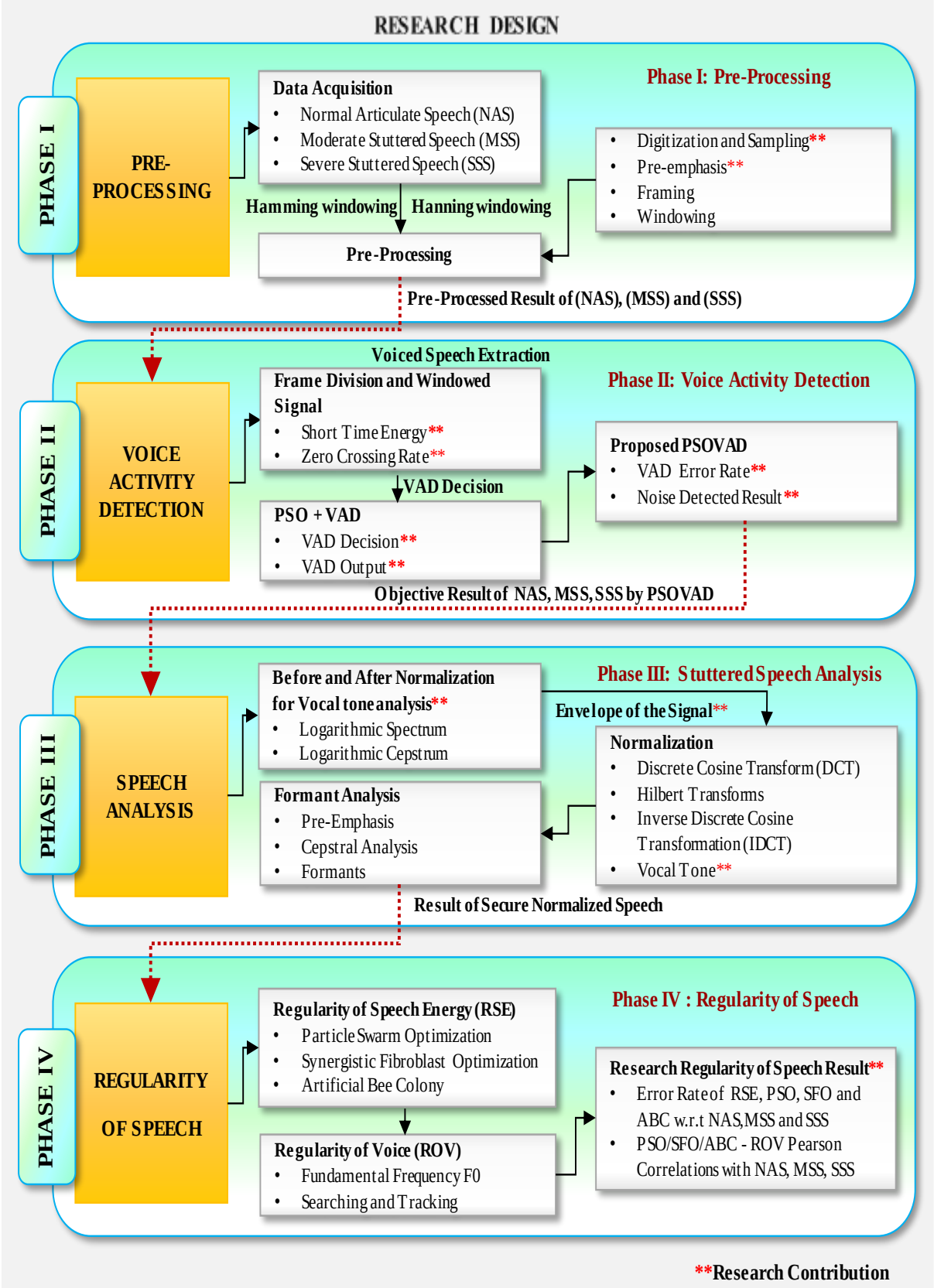

Figure 4.1. Proposed Research Design

CHAPTER I : DATA ACQUISITION

Initially, 15 people (male) with developmental stuttering were identified with the help of a speech pathologist and also people speaking Tamil language very fluently, that is, Normal Articulate Speech (NAS) (5 males) were selected to create the speech database. Out of 15 people, 7 participants were Moderate Stutter Speech (MSS), and 3 participants were Severe Stutter Speech (SSS), which has been categorized based on the SSI-3 (Stuttering Severity Instrument) evaluation [8] [12].

The real-time data was recorded with Sony ICD- UX560F Digital Voice Recorder. In this, fifty simple Tamil sentences were developed with the help of speech pathologists and a passage from the sixth standard Tamil school book, namely “Saathanai Penmani Marie Curie” (meaning Achievement Women Marie Curie), given for the oral reading to generate the database. There are 500 samples recorded concerning the stuttering speaker (Normal Articulate Speech (NAS), Moderate Stutter Speech (MSS), and Severely Stutter Speech (SSS)). The simple and complex sentences in the Tamil language recorded a duration of 60 seconds to 7 minutes. Around 664 × 5 × 15 Tamil speech signals were acquired, i.e., the speakers uttered 664 words and were further allowed to repeat the entire data 5-times with different speakers, including a regular voice of the total recorded words of 49,800. Among that, 2700 words (NAS [900], MSS [900], and SSS [900]) were considered for the proposed research work depending on the syllable repetition, utterance delay and silent duration of stuttered speakers. The recorded Tamil speech files were stored in the .wav file format with a sampling frequency of 16000 kHz.

CHAPTER II : PRE-PROCESSING

The digitized speech exemplifies a speech signal as a continuous distribution of numbers sampled by calculating its amplitude at that little point in time. Generally, the speech signal is pre-emphasized before any other type of processing because this process is used to raise the high-frequency spectral components in the speech signal. Hence the pre-emphasized speech signal is divided by an adjacent number of frames [11]. In the case of tiny frame size, the speech samples were insufficient to compute a reliable spectral estimation and reduce the frequency resolution. The missing data were avoided by using the overlapping frames so that the current frame envelopes some of the information from the earlier frame. Splitting speech signals into a small number of frames lead to distorted

spectral information in segmented speech. The purpose of a windowing function is to minimize the impact of spectral artefacts obtained from the framing process to avoid a discontinuity at each edge of frames. Applying the window function reduces the variation that occurs among the speech frames. Also, it improves the frequency spectrum of the speech signals, which is an exceptional aid in analysing the speech severity rate of stutter speakers [10]. A windowing function was performed with the Hamming window and Hanning window. The hamming window reduces the overlapping by 95% in many cases since it has excellent resolution and minimizes spectral leakage. The Stutter Speech Severity Rate (SSR) and Data Acquisition Rate (DAR) of ordinary and stutter speech before and after pre-emphasis were computed.

Results of Pre-processing:

The experimental results produced by the hamming window outperform well with the Hanning window as the Hamming window removes the nearest side lobe better than the Hanning window.

CHAPTER III : VOICE ACTIVITY DETECTION

The significant step after the pre-processing is a reliable partition of voiced and unvoiced speech segmentation, successfully achieved using the Voice Activity Detection (VAD) method. The VAD algorithm is used to discriminate the presence or absence of speech, silence, or even any other background noise in the speech samples. The detection process is not as trivial as it appears because the noise accumulated in the background of a speech signal reduces the effectiveness of classifiers. The working mechanism behind the VAD depends on any sequence of public speech properties like periodicity, spectrum, and temporal energy variation [2]. Zero crossings and energy are significant components in voice activity detection to segment the speech region into voiced, unvoiced, and silence. Hence each frame's features have been extracted by using typical parameters like Short Time Energy (STE), Zero Crossing Rate (ZCR), Normalized Auto Correlation (NAC), and Auto-Correlation Periodicity. The extracted features assist in taking the Voice Activity Detection decision. Initially, the energy that detects the voiced speech obtained from the classical method,

- If the measured value exceeds the threshold, it is considered active speech (VAD=1).
- Otherwise, if the measured value does not exceed the limit, it is treated as a no-speech activity (VAD=0).

The speech signal temporal feature has been extracted from the energy and zero-crossing rate, which is initially optimized by Particle Swarm Optimization (PSO). PSO is proposed in this research to improve active speech signal detection efficiently. A short-time energy threshold is computed from each incoming frame. The conventional VAD and PSOVAD were evaluated using four standard objective parameters like Front End Clipping (FEC), Mid-Speech Clipping (MSC), Over Hang (OVER), and Noise Detected as Speech (NDS).

Results of Voice Activity Detection:

The experimental results suggest that the proposed PSOVAD outperforms competing conventional VAD providing accurate detection of the voiced and unvoiced region. The proposed PSOVAD method minimizes the error rate. Compared to the traditional approach, PSOVAD enhances the detection of active speech and minimizes the error rate. The overall error rate of the proposed PSO-based VAD is decreased to 79.7% compared to the conventional method. The performance of the PSOVAD method has been evaluated based on the objective parameters under noise environments for NAS, MSS, and SSS.

Publications:

1. Manjutha, M., Subashini. P., (2022), "Statistical Model-Based Tamil Stuttered Speech Segmentation Using Voice Activity Detection", *Journal of Positive School Psychology*, Vol. 6, No. 8, pp. 1461–1473. ISSN 2717-7564 (UGC CARE: Group II – Scopus Indexed)
2. M. Manjutha, P. Subashini, M. Krishnaveni, (2019), "Vocal Tone Analysis for Identification of Stuttering Levels based on Tamil Syllable", *International Journal of Recent Technology and Engineering (IJRTE)*, Vol.8, No.2, ISSN: 2277-3878, pp. 4472-4483. DOI: 10.35940/ijrte.B3211.078219. (UGC CARE: Group II – Scopus Indexed)

CHAPTER IV : SPEECH ANALYSIS

People affected by stuttering have great difficulty communicating with others when their vocal apparatus fails. Human speech production consists of two separate processes in which one produces initial sound, and another modifies it. The larynx generates a speech sound whose spectrum has many different frequencies [5]. The proposed work consists of two significant analyses, which analyses the real-time Tamil stuttered speech [1].

1. Vocal Tone Analysis
2. Formant Analysis

The vocal folds produce the sound, but the tone is modified by manipulating the different structures of the vocal tract. The proposed vocal tone analysis consists of data acquisition, logarithmic spectrum before normalization, cepstrum before normalization, an envelope of the signal spectrum, normalization, logarithmic spectrum after normalization, cepstrum after normalization and vocal tone. From the vocal tone, spectrum, and cepstrum of the stuttered speech signal computed. The obtained result was further processed to identify the envelope of the signal. The signal envelope was normalized by spectrum and cepstrum since the vowels and fricative produced by ordinary people and stutter people have different spectral shapes of the vocal tract. In the proposed work, normalization was applied to predict the vocal tone of the people affected by stuttering.

Due to the frequency shaping of the speech signal, the formant frequencies were derived from the vocal folds by the vocal tract. That includes the nasal tract, tongue, teeth, lips, palate, etc. The particular configuration of the above organs (articulators) for every phoneme creates resonances at specific frequencies called formants [6][7]. Hence formants exist for both voiced and unvoiced sounds. In this research, three formant frequencies, such as Formant one (F1), Formant two (F2), and Formant three (F3) computed for the perception of fluent and disfluent speakers.

Results of Speech Analysis:

The experimental outcomes of vocal tone identify that people affected by severe stutter have low vocal tone than moderate stutter and ordinary speech, which is represented as subjective evaluation in three categories such as NAS, MSS, and SSS.

The formant analysis was done by evaluating formant frequency fluctuation that results varies depending on the Tamil syllables for NAS, MSS, and SSS.

Publications:

1. Manjutha, M., Krishnaveni, K., and Subashini, P. (2020), "Acoustic Analysis of Formant Frequency Variation in Tamil Stuttered Speech," *Journal of Advance Research in Dynamical and Control Systems*, vol.12, no. 07, 2020, pp.2934-2944, DOI:10.5373/JARDCS/V12SP7/20202438 (UGC CARE: Group II – Scopus Indexed)

2. M. Manjutha, P. Subashini, M. Krishnaveni, (2019), "Vocal Tone Analysis for Identification of Stuttering Levels based on Tamil Syllable", International Journal of Recent Technology and Engineering (IJRTE), Vol.8, No.2, ISSN: 2277-3878, pp. 4472-4483. DOI: 10.35940/ijrte.B3211.078219. (UGC CARE: Group II – Scopus

CHAPTER V : REGULARITY OF SPEECH

The speech signal tempo will vary with each individual where the specific fluctuation in the velocity of stutter speech is typical, and it is due to the intervals in the speech rate, which has a significant difference in regular and stutter people [9]. The acoustic measurement has been evaluated to distinguish the regularity of Tamil fluent and dysfluent speech by introducing the following objective parameters such as,

1. The Regularity of Speech Energy (RSE)

2. The Regularity of Voice (ROV)

The Tamil-speaking dataset computed the Regularity of Speech Energy (RSE) for regular, moderate, and severe stutter speech. The RSE was done based on the energy threshold obtained during the irregular release of energy, which henceforth was analysed

using optimal thresholding based on Particle Swarm Optimization (PSO), Synergistic Fibroblast Optimization (SFO) and Artificial Bee Colony (ABC) techniques.

The experimental results of RSE have been evaluated with statistical measures such as mean, standard deviation, Root Mean Square Error (RMSE), and Signal Noise Ratio (SNR). The RSE results suggest that stuttered speaker's signal releases low energy compared to the typical speaker, where the optimal threshold energy enhances the detection of hidden speech energy.

The methods of F0 estimation are divided into three categories: based on temporal dynamics of the signal or time-domain based on the frequency structure, or frequency-domain, and hybrid approaches. The regularity of voice or fundamental frequency (F0) estimation consists of three significant steps.

1. Pre-processing (signal filtering, splitting it into frames)

2. Searching for possible values for F0 (candidates)

3. Tracking: the choice of the most probable F0 trajectory (which is essential for each moment that has several simultaneously competing candidates)

The fundamental frequency (F0) is computed from the short-time windowed function. The obtained F0 values were stored in the accumulator. The limit of the accumulator value (threshold) is set individually for every speech signal as the median value of pitch period F0. The threshold value is further optimized with Particle Swam Optimization (PSO), Synergistic Fibroblast Optimization (SFO), and Artificial Bee Colony (ABC). Since the existence of fundamental frequency only makes sense inside voiced frames, all the outcomes related to F0 are confined to voiced regions. Statistical measures such as mean, standard deviation, standard error, and Pearson correlation were performed to evaluate the experimental analysis on ROV.

Results of Speech Analysis

The proposed PSO-RSE enhances the spectrum energy more efficiently for fluent and dysfluent speech. The proposed work has been evaluated by using Root Mean Square Error Rate (RMSE) and Signal to Noise Ratio (SNR) for the conventional method (RSE), PSO-RSE, SFO-RSE, and ABC-RSE in which PSO-RSE reduces the error rate more effectively. Compared to the traditional approach, PSO-RSE enhances the energy at

76.11% and the conventional SFO-RSE and ABC-RSE at 69.09%, 75.04% and 75.69%, respectively.

Also, the Regularity of Voice has evaluated that PSO-ROV identifies the fundamental frequency efficiently at 3.87% compared to other SFO-ROV and ABC-ROV methods of 3.31% and 3.72%, respectively.

Publications:

1. Manjutha M, P. Subashini, M. Krishnaveni, "Analysis on Regularity of Speech Energy based on Optimal Thresholding for Tamil Stuttering Dataset", In the proceeding of *5th IEEE International Smart Cities Conference 2019*, October 14-17 2019, Casablanca, Morocco, ISBN: 978-1-7281- 0846-9 pp.177-183.

5. RESEARCH CONTRIBUTIONS

The research aims to efficiently analyze the Tamil stuttered speech by introducing computational intelligence algorithms that associate the speech pathologist with dysfluent speakers to develop an efficient diagnosis.

Chapter I: The real-time Tamil stuttered speech database has been created

Chapter II: In pre-processing, the Hamming window is proposed to reduce the signal overlapping.

Chapter III: PSO-VAD has been introduced to enhance the detection of voiced, unvoiced and silence exist in the NAS, MSS and SSS

Chapter IV: The vocal tone and formant analysis proposed to identify the frequency variation among fluent and disfluent speech.

Chapter V: The regularity of speech energy and the regularity of voice has been computed for NAS, MSS and SSS by introducing Particle Swam Optimization (PSO), Synergistic Fibroblast Optimization (SFO) and Artificial Bee Colony (ABC).

6. RESEARCH HIGHLIGHTS

The speakers' overall performance results in stuttered speakers having difficulty reproducing the spectra as ordinary speakers. Depending upon the types of a syllable in the Tamil language, the appropriate recovery time required for a severely stuttered speaker outperforms similar of ordinary people. Compared to the conventional VAD, the PSOVAD method efficiently enhances the detection of voiced and unvoiced parts of the speech signal uttered by NAS, MSS, and SSS. The vocal tone analysis results that the severe stuttered speakers expressed more low-frequency variations and fluctuations in the time interval during their utterances. Formant Analysis detects the changes in the frequency spectrum of NAS, MSS, and SSS for F1 and F2 and F3 of different Tamil Syllables. The energy of fluent and dysfluent speech is computed to identify the dissimilarity in which the speech energy of PSO-RSE strengthens the conventional RSE. Also, PSO-RSE reduces the Mean Square Error (MSE) compared to traditional RSE, SFO-RSE, and ABC-RSE. Tracking of voice regularity or fundamental frequency was uncomplicated by using PSO-ROV, which yields a significant differentiation in the vocal fold vibration of fluent and dysfluent speech, which reduces the standard error.

7. REFERENCES

[1] Andrzej Czyzewski, Andrzej Kaczmarek and Bozena Kostek (2003), „Intelligent Processing of Stuttered Speech“, Journal of Intelligent Information Systems, Vol.21, No.2, pp.143–171.

[2] B. Adapa, Rajesh Bachu,S. Kopparthi and Buket D. Barkana,” Voiced/Unvoiced Decision for Speech Signals Based on Zero-Crossing Rate and Energy”, Advanced Techniques in Computing Sciences and Software Engineering, USA, 2009,pp.279-282. DOI:10.1007/978-90-481-3660-5_47.

[3] Banurekha Velayutham, Boopathi Kangusamy and Sanjay Mehendale, “Prevalence of disability in Tamil Nadu, India”, The National Medical Journal of India, vol.30, no. 3, pp.125-130, 2017.

[4] Census of India 2011, “Office of the Registrar General and Census Commissioner”, New Delhi, India, http://www.censusindia gov.in Accessed on 27-12-2013.

[5] E. Szabelska, A Kruczyńska (2013), "Computer-based speech analysis in stutter", Applied Computer Science, vol. 9, no.2, pp.34-42.

[6] Eichhorn, J., Kent, R., Austin, D., and Vorperian, H. K. (2018). “Effects of aging on vocal fundamental frequency and vowel formants in men and women”, Journal of Voice,vol.32,no.5, pp. 644.e1-644.e9, http://dx.doi.org/10.1016/ j.jvoice.2017.08.003

[7] Elham Masumi, Zohre Arani Kashani, Nafise Hassanpour and Mohammad Kamali (2015), "The Effect of Syllable Structure on the Frequency of Disfluencies in Adults with Stuttering", Middle East Journal of Rehabilitation and Health, vol.2, no.2, e26436, pp.1-5.

[8] Leanne Jani, Mark Huckvale and Peter Howell (2013), "Procedures used for assessment of stuttering frequency and stuttering duration", Clinical Linguistics & Phonetics, vol.27, no.12, pp.853–861.

[9] Leona Polyanskaya, Arthur G. Samuel and Mikhail Ordin, “Regularity in speech rhythm as a social coalition signal”, Annals of the New York Academy of Sciences, pp.1-13, 2019, DOI:10.1111/nyas.14193

[10] Miragaia, R., Fernández, F., Reis, G., & Inácio, T. (2021). Evolving a Multi-Classifier System for Multi-Pitch Estimation of Piano Music and Beyond: An Application of Cartesian Genetic Programming. Applied Sciences, 11(2902), 1–27. https://doi.org/10.3390/app11072902

[11] R.N.Mutagi, “Understanding the sampling process”, September 2004, pp.38-48.

[12] Riley, G. D. (1972). A Stuttering Severity Instrument for Children and Adults. “Journal of Speech and Hearing Disorders”, 37(3), 314–322. https://doi.org/10.1044/jshd.3703.314

[13] Shreya Pareek. (2015, August). I Stammer and It Is Completely OK – This Team Aims to Support 1.25 Crore ‘Stammerers’ in India. The Better India. https://www.thebetterindia.com/31784/stammering-in-india-tisa/

[14] The Stuttering Foundation and Frank Wolf, "Daily news report analysis", 2010, [Online]. Available: [http://www.stutteringhelp.org/].

[15] The Stuttering Foundation and Frank Wolf, Stuttering Gets Congress’ Attention. (2014, November). Summer 2014 Newsletter. http://www.stutteringhelp.org/ content/stuttering-gets-congress%E2%80%99-attention

8. PUBLICATIONS

Journals

1. Manjutha, M., Subashini. P., (2022), "Statistical Model-Based Tamil Stuttered Speech Segmentation Using Voice Activity Detection", *Journal of Positive School Psychology*, Vol. 6, No. 8, pp. 1461–1473. ISSN 2717-7564 (UGC CARE: Group II – Scopus Indexed)

2. Manjutha, M., Subashini. P., (2022), "Survey on optimization algorithms in speech processing", *International Journal of Health Sciences*, Vol. 6, No. S5, pp. 2997–3017. ISSN 2550-6978, DOI: https://doi.org/10.53730/ijhs.v6nS5.9307 (UGC CARE: Group II – Scopus Indexed)

3. Manjutha, M., Krishnaveni, K., and Subashini, P. (2020), "Acoustic Analysis of Formant Frequency Variation in Tamil Stuttered Speech," *Journal of Advance Research in Dynamical and Control Systems*, vol.12, no. 07, 2020, pp.2934-2944, DOI:10.5373/JARDCS/V12SP7/20202438 (UGC CARE: Group II – Scopus Indexed)

4. M. Manjutha, P. Subashini, M. Krishnaveni, (2019), "Vocal Tone Analysis for Identification of Stuttering Levels based on Tamil Syllable", *International Journal of Recent Technology and Engineering (IJRTE)*, Vol.8, No.2, ISSN: 2277-3878, pp. 4472-4483. DOI: 10.35940/ijrte.B3211.078219. (UGC CARE: Group II – Scopus Indexed)

5. M. Manjutha, P. Subashini, (2017), "Particle Swarm Optimization based Voice Activity Detection for Stuttered Tamil Speech", *International Journal of Computer Engineering and Applications*, Vol.9, No.9, ISSN: 2321-3469, pp. 1-14.

6. Manjutha M, Gracy J, Dr. P Subashini, Dr. M Krishnaveni (2017), "Automated Speech Recognition System – A Literature Review", *International Journal of Engineering Trends and Applications (IJETA)* Mar - Apr 2017, Published by Eighth Sense Research Group. ISSN: 2393-9516 ,V4(2), Issue(2), pp.42-49. (Extended Version).

Conferences

1. Manjutha M, P. Subashini, M. Krishnaveni, "Analysis on Regularity of Speech Energy based on Optimal Thresholding for Tamil Stuttering Dataset", In the proceeding of *5th IEEE International Smart Cities Conference 2019*, October 14-17 2019, Casablanca, Morocco, ISBN: 978-1-7281- 0846-9 pp.177-183.

2. M. Manjutha, J. Gracy, Dr. P. Subashini and Dr. M. Krishnaveni, "Automated Speech Recognition System — A Literature Review", In the proceeding of *5th National Conference on Computational Methods, Communication Techniques and Informatics 2017*, The Gandhigram Rural Institute Deemed University, Dindigul, Tamilnadu, India, ISBN: 978-81-933316-1-3, pp.205- 212.